Visualizing: ggplot2

AW

2022-07-01

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
## from CRAN:
install.packages("tidyverse") ## do this once only to install the package on your computer.

library(tidyverse) ## do this every time you restart R and need it
```

Load Data

The following R code chunk import the data needed. It has the only option `message=FALSE` so that this R code chunk will be evaluate but the output and message will be suppressed.

```
#National Parks in California
ca <- read_csv("https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/ca.csv")

#Acadia National Park
acadia <- read_csv("https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/acad

#Southeast US National Parks
se <- read_csv("https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/se.csv")

#2016 Visitation for all Pacific West National Parks
visit_16 <- read_csv("https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/vi

#All Nationally designated sites in Massachusetts
mass <- read_csv("https://raw.githubusercontent.com/OHI-Science/data-science-training/master/data/mass.
```

Plotting with ggplot2

`ggplot2` is a plotting package that makes it simple to create complex plots from data in a data frame.
`ggplot` likes data in the 'long' format: i.e., a column for every dimension, and a row for every observation. Well structured data will save you lots of time when making figures with `ggplot`.
`ggplot` graphics are built step by step by adding new elements. Adding layers in this fashion allows for extensive flexibility and customization of plots.

National Park visitation dataset `ca` example This dataframe is already in a long format where all rows are an observation and all columns are variables. Among the variables in `ca` are:

1. `region`, US region where park is located.
2. `visitors`, the annual visitation for each year

To build a ggplot, we need to:

- use the `ggplot()` function and bind the plot to a specific data frame using the data argument
- add `geoms` – graphical representation of the data in the plot (points, lines, bars). `ggplot2` offers many different `geoms`; we will use some common ones today, including: * `geom_point()` for scatter plots, dot plots, etc. * `geom_bar()` for bar charts * `geom_line()` for trend lines, time-series, etc.
  To add a `geom` to the plot use `+` operator. Because we have two continuous variables, let's use `geom_point()` first and then assign x and y aesthetics (`aes`):

```
ggplot(data = ca) +
  geom_point(aes(x = year, y = visitors))
```